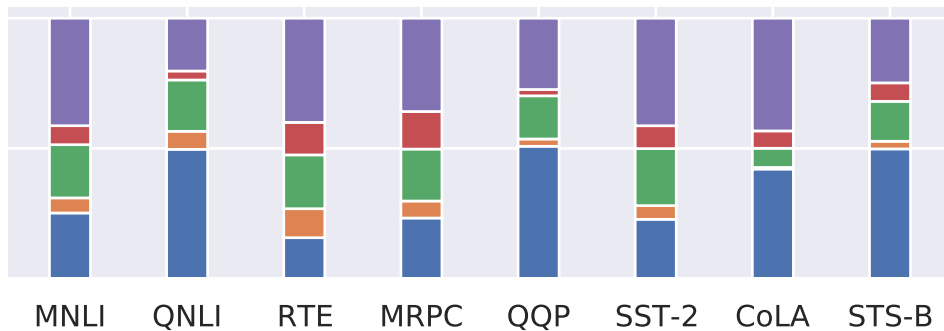Attention Patterns

Weight Normed Attention Patterns

MNLI QNLI RTE MRPC QQP SST-2 CoLA STS-B

Vertical + Diagonal   Vertical   Diagonal
Block   Heterogeneous